Décision Bayésienne

Thierry Chateau

Pascal Institute

2017

Plan

Principe de la décision Bayesienne

- La théorie de la décision de Bayesienne constitue une approche fondamentale de la RdF
- c'est une méthode de reconnaissance stochastique
- Elle suppose que le problème peut être entièrement spécifié en termes de probabilités
- Sous ces hypothèses, la décision Bayesienne peut être considérée comme optimale.

Exemple introductif

- Exemple d'une entreprise de sciage de troncs d'arbres qui ne traîte que des frênes et des bouleaux.
- Etat d'une planche = "frêne" (classe ω_1) ou bouleau (classe ω_2).
- On définit une variable aléatoire X qui ne peut prendre que 2 valeurs (ω_1 ou ω_2).

Exemple introductif

Probabilité à priori d'une classe

- On suppose que l'on connaît les proportions de planches en sortie car on connaît les quantités de frêne et de bouleau en entrée (par exemple $2/3$ de frêne et $1/3$ de bouleau.
- **Question** : Sans aucune information, comment décider la classe d'une planche quelconque ?
- **Réponse** : On décidera que c'est du frêne (minimisation de la probabilité de se tromper)
- En fait, on dispose d'une information importante (**les probabilités a priori**):
  - $p(\omega_1) = 2/3$
  - $p(\omega_2) = 1/3$

Exemple introductif

Probabilité à priori d'une classe (suite)

- Lorsque aucune information à priori n'est connue, des probabilités équiprobables sont choisies
- La plupart du temps, les probabilités a priori sont fixées par apprentissage.

Règle de Bayes

- Soit $\{\omega_1, \omega_2, ...\omega_c\}$ un ensemble de c classes et $\mathbf{x}$ un vecteur de caractéristiques.
- Pour chaque classe ω_i on suppose connaître :
  - $P(\omega_i)$: la probabilité a priori de cette classe,
  - $p(\mathbf{x}|\omega_i)$: la densité de probabilité de $\mathbf{x}$ conditionnée par cette classe

Règle de Bayes

- La règle de Bayes permet de calculer la probabilité d'une classe a posteriori, c'est-à-dire conditionnée par l'observation de x, soit

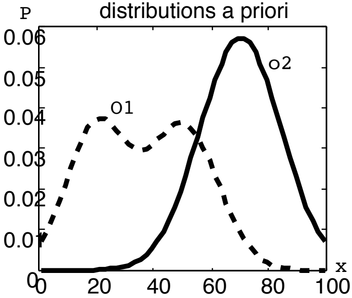
$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}$$

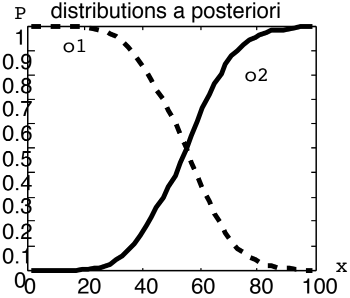
avec :

$$p(\mathbf{x}) = \sum_i (p(\mathbf{x}|\omega_i).P(\omega_i))$$

Illustration Règle de Bayes

2 classes

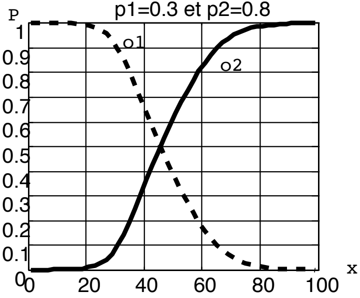


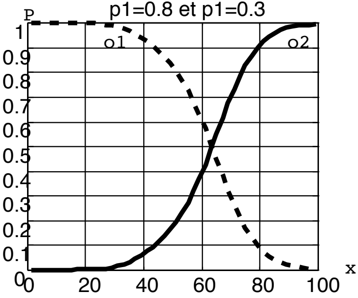


$$P(\omega_1) = P(\omega_2) = 0.5$$

Modif $P(\omega_i)$

2 classes





1. A gauche $P(\omega_1) = 0.3,\ P(\omega_2) = 0.7$
2. A droite $P(\omega_1) = 0.7,\ P(\omega_2) = 0.3$

Probabilité d'erreur

Soit un vecteur observé $\mathbf{x}$ et une décision prise $\delta(\mathbf{x}) = \omega_i$, le probabilité d'erreur associée à la décision est :

$$P(\text{error}|\mathbf{x}) = \sum_{j \neq i} P(\omega_j|\mathbf{x}) = 1 - P(\omega_i|\mathbf{x})$$

La probabilité d'erreur globale associée au système est :

$$P(\text{errorglob}|\mathbf{x}) = \int_{-\infty}^{\infty} P(\text{error}|\mathbf{x}).P(\mathbf{x})dx$$

Décision optimale

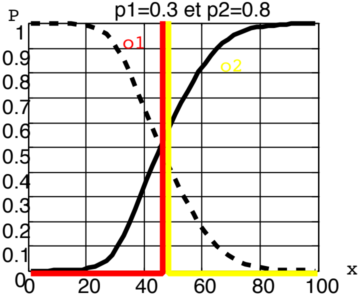
La décision optimale (au sens de la probabilité d'erreur) consiste donc à choisir $\delta(\mathbf{x}) = \omega_i$ telle que $p(\omega_i|x)$ soit maximale :

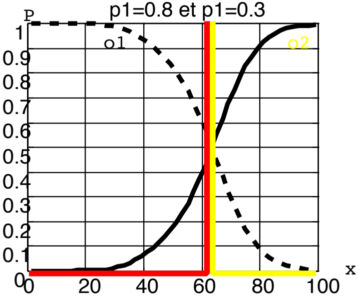
$$P(\omega_i|\mathbf{x}) \geq P(\omega_j|\mathbf{x}) \forall j$$

ou encore :

$$p(\mathbf{x}|\omega_i).P(\omega_i) \geq p(\mathbf{x}|\omega_j).P(\omega_j) \forall j$$

2 classes (régions de décision, frontières de décision)





1. A gauche $P(\omega_1) = 0.3, \; P(\omega_2) = 0.7$
2. A droite $P(\omega_1) = 0.7, \; P(\omega_2) = 0.3$

Notion de cout et de risque

- Soit $\{\delta_1, \delta_2, ..\delta_d\}$, l'ensemble des décisions possibles, δ_i correspondant à $\delta(\mathbf{x}) = \omega_i$
- Soit $\lambda(\delta_i|\omega_j)$ le coût engendré par la décision δ_i lorsque l'objet appartient effectivement à la classe ω_j
- la probabililé d'erreur correspond au cas particulier où :

$$\lambda(\delta_i, \omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \tag{1}$$

Risque

- Le risque associé à la décision δ_i (risque conditionnel) est:

$$R(\delta_i|\mathbf{x}) = \sum_j \lambda(\delta_i|\omega_j)P(\omega_j|\mathbf{x})$$

- Le risque global est déterminé par :

$$R = \int_{R^n} R(\delta(\mathbf{x})|\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

- Minimiser le risque global revient à prendre, pour chaque valeur de $\mathbf{x}$, la décision qui minimise le risque conditionnel.

Cas de 2 classes

- on pose $\lambda_{ij} = \lambda(\delta_i|\omega_j)$ (cout de la décision δ_i alors que la décision est ω_j)

$$R(\delta_1|\mathbf{x}) = \lambda_{11} p(\omega_1|\mathbf{x}) + \lambda_{12} p(\omega_2|\mathbf{x})$$

$$R(\delta_2|\mathbf{x}) = \lambda_{21} p(\omega_1|\mathbf{x}) + \lambda_{22} p(\omega_2|\mathbf{x})$$

avec : $\lambda_{11} < \lambda_{12}$ et $\lambda_{21} < \lambda_{22}$ car la décision qui correspond à la vérité doit couter la moins cher :

$$\omega_1 \text{ si } R(\delta_1|\mathbf{x}) < R(\delta_2|\mathbf{x})$$

Cas de 2 classes

- d'où
$$(\lambda_{21} - \lambda_{11})P(\omega_1|\mathbf{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2|\mathbf{x})$$
- soit
$$(\lambda_{21} - \lambda_{11})P(\mathbf{x}|\omega_1)P(\omega_1) > (\lambda_{12} - \lambda_{22}p(\mathbf{x}|\omega_1)P(\omega_1)$$
- finalement on décide ω_1 si :
$$\frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}\frac{P(\omega_2)}{P(\omega_1)}$$
- Ce rapport est **le rapport de vraissemblance**

Classification par taux d'erreur minimum

- Fonction de coût symétrique : $\lambda_{ij} = 0$ si $i = j$ et $\lambda_{ij} = 1$ si $i \neq j$.
- Rappel du risque :

$$R(\delta_i|\mathbf{x}) = \sum_j \lambda(\delta_i|\omega_j)P(\omega_j|\mathbf{x})$$

$$R(\delta_i|\mathbf{x}) = \sum_{j \neq i} P(\omega_j|\mathbf{x}) = 1 - P(\omega_i|\mathbf{x})$$

- Minimiser le risque revient à maximiser les probabilités à postériori

Fonctions discriminantes

- Pour définir une règle de décision, on utilise une fonction discriminante, définie par $g_i(x)$, $i = 1, ..s$ (s=nb de classes)
- les s fonctions discriminantes sont telles que une forme $\mathbf{x}$ est classée dans ω_i si $g_i(\mathbf{x}) > g_j(\mathbf{x})\ \forall j \neq i$

Fonctions discriminantes

Plusieurs fonctions discriminantes peuvent être définies :

- $g_i(\mathbf{x}) = -R(\delta_i|\mathbf{x})$
- $g_i(\mathbf{x}) = p(\omega_i|\mathbf{x})$
- $g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)$
- $f(g_i(\mathbf{x}))$ si f est une fonction monotone croissante et g_i est une fonction discriminante.

Fonctions discriminantes

On trouve, de manière usuelle, les fonction suivantes :

- $g_i(\mathbf{x}) = p(\omega_i|\mathbf{x})$
- $g_i(\mathbf{x}) = \dfrac{p(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_i (p(\mathbf{x}|\omega_i).P(\omega_i))}$
- $g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)$
- $g_i(\mathbf{x}) = log(p(\mathbf{x}|\omega_i)) + log(p(\omega_i))$